**What is Machine Learning?**

Anna Wong, Kyrrah Nork, Joel Aguiar

ITC 134

February 13, 2017

# Abstract

The purpose of this paper is to find out how machine learning works.  We have looked into the history and researched what tools are currently being used.  We discuss what ways machine learning is used in one’s everyday life and the implications it may have on society; and in conclusion, why we, as developers and programmers, should be familiar with this technological tool.

# Introduction

The idea of teaching a machine to learn from experience goes back many decades. Beginning in 1952 an employee at IBM, Arthur Samuel, wrote the first computer learning program. This was a game of checkers, and the computer improved at the game the more it played**.** Over the intervening decades massive leaps in machine learning have been made. Currently many companies whose websites are utilized millions of times daily, make use of these advancements in learning-algorithms to do what they hope is make our lives simpler and easier. Despite the good intentions behind the use of these tools this is not always the case. In this paper we discuss the concept of machine learning broadly and provide some specific examples of both the negative and positive effects this technology is having.

# What is Machine Learning? A Brief History

According to Wikipedia, “machine learning is the subfield of computer science that gives computers the ability to learn without being explicitly programmed” (Machine Learning 2017). The history of machine learning goes back much farther than one might think. As far back as 1950 when Alan Turing created the “Turing Test”, a test to determine if a computer has real intelligence (Marr 2016), the scientific community was considering whether machines could not only think, but learn as well. This concept was proven to be a reality when in 1952 “Arthur Samuel wrote the first computer learning program. The program was the game of checkers, and the computer improved at the game the more it played, studying which moves made up winning strategies and incorporating those moves into its program.” (Marr 2016). Jumping quickly through the intervening decades we have:

* 1957 - Frank Rosenblatt designed the first neural network for computers, the perceptron.
* 1967 - The “nearest neighbor” algorithm was written, allowing computers to begin using very basic pattern recognition.
* 1979 - Students at Stanford University invent the “Stanford Cart” which can navigate obstacles in a room on its own.
* 1997 - IBM’s Deep Blue beats the world champion at chess.
* 2010 - The Microsoft Kinect can track 20 human features at a rate of 30 times per second, allowing people to interact with the computer via movements and gestures.
* 2011 - Google Brain is developed, and its deep neural network can learn to discover and categorize objects much the way a cat does.
* 2014 - Facebook develops DeepFace, a software algorithm that is able to recognize or verify individuals on photos to the same level as humans can.
* 2015 - Amazon launches its own machine learning platform.
* 2015 - Microsoft creates the Distributed Machine Learning Toolkit, which enables the efficient distribution of machine learning problems across multiple computers.
* 2015 - Over 3,000 AI and Robotics researchers, endorsed by Stephen Hawking, Elon Musk and Steve Wozniak (among many others), sign an open letter warning of the danger of autonomous weapons which select and engage targets without human intervention.
* 2016 - Google’s artificial intelligence algorithm beats a professional player at the Chinese board game Go, which is considered the world’s most complex board game and is many times harder than chess. The AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Go competition. (Marr 2016)

This list is by no means all-inclusive of the vast number of advances in machine learning.

# How is Machine Learning Being Used Today?

Stitchfix is a service that helps busy people take some of the guess work out of shopping for their wardrobe.  Stitchfix uses “recommendation algorithms “to help pick out clothes to send to their clients as a monthly service.  The client takes a survey about themselves and through that profile the company predicts merchandise that the clients will be interested in.  They then send five pieces to the client's home for approval and they keep what they like and send the rest back.  They have several teams equaling 70 people that constantly work on this side of the business.  These teams include:

* Client Algorithms whose mission is to understand the client and predict their demands.
* Style Algorithms is where the fashion is quantified and shipped.
* Merch Algorithms is where the magic of fashion, inventory and logistics happen. (MultiThreaded n.d.)

Another company that utilizes machine learning that most people worldwide are familiar with is Google. Google uses what are referred to as deep neural networks of “hardware and software to approximate the web of neurons in the human brain” (Metz 2016). These networks are able to sift through vast amounts of data, analyze it, and learn to do tasks such as “identifying photos, recognizing commands spoken into a smartphone, and, as it turns out, responding to Internet search queries” (Metz 2016). This ability to learn allows these networks to perform these tasks at a far greater speed and better than humans, which in turn makes for a vast increase in the amount of data processed.

This has led into the use of “a deep learning system called RankBrain that helps generate responses to search queries. As of October, RankBrain played a role in ‘a very large fraction’ of the millions of queries that go through the search engine with each passing second” (Metz 2016). Some concerns remain regarding these neural nets and the fact of the matter is that even the experts do not completely understand how they work. They do know that:

“if you feed enough photos of a platypus into a neural net, it can learn to identify a platypus. If you show it enough computer malware code, it can learn to recognize a virus. If you give it enough raw language—words or phrases that people might type into a search engine—it can learn to understand search queries and help respond to them” (Metz 2016).

Google began by using algorithms to drive their search engine that would automatically generate a response to queries. These were basically rules that were difficult or couldn’t be changed. With the incorporation of machine learning and AI Google believes this is the future. The use of neural nets may not achieve the results desired—at first, but over time as the machine learns the results returned improve in accuracy as well as intuiting what the searcher is actually looking for even when they might not know what to type into the search.

# What are some of the tools?

There is a myriad of tools for using machine learning techniques. These tools use supervised learning, unsupervised learning, and reinforcement learning to teach the machine the desired skill. (Ramasubramanian 2016) Recently, there has been a growing movement to make this technology accessible to more researchers and developers which has resulted in open source machine learning software. Some of the machine learning tools include Scikit-Learn, TensorFlow, and CafeOnSpark.

* CafeOnSpark is an open source program that specializes in speech recognition or photo and video content recognition.
* DeepLearning4j is a deep learning library.  It is written in java, scala, CUDA, C, C++. Released under Apache license 2.0 It's the first commercial grade open source library.  It can import neuron net models. Used for fraud detection and network intrusion detection, face recognition, voice search.  Its tools configure neural networks. Neural networks are algorithms that recognize patterns and are then put into classifications called clustering. (DeepLearning4J n.d.)
* Scikit-Learn is a powerful tool that is used by big companies like Spotify and Evernote. It is ‘Machine Learning in Python’, so it is very easy to use as long as one is familiar with the Python programming language. One can use it to achieve most of the common machine learning tasks including classification, regression, clustering, model selection, dimensionality reduction, and preprocessing. Scikit-Learn comes with a lot of documentation and dozens of example projects with the corresponding python code. This makes it a very accessible tool.
* TensorFlow is an “open-source software library for Machine Intelligence” (Corporate n.d.). It was first developed by Google employees working for the Machine Intelligence research organization. Like Scikit-Learn you can use Python code to control it. It also comes with dozens of tutorials that make it easy to get started. Not surprisingly, given apps developed by Google, TensorFlow can be used for image processing and language and sequence processing. Using a deep convolutional neural network, TensorFlow can help with recognizing objects, animals, and people in pictures.

# Why Does Machine Learning Matter for Programmers and Developers?

Machine learning can be a double edged sword for developers. According to the InfoWorld website,” Cloud-based machine learning tools can act as a way for developers to dip their toes into the possibilities that machine learning creates and can offer novel functionality. When used incorrectly, however, these tools garner poor results, which can be frustrating for users” (Gray 2015). To avoid some of the pitfalls associated with machine learning programmers should keep at least the following three things in mind. First, be sure to use the learning method that works best for the problem you are trying to solve. Second, models for predictive machine learning are only as good as the information they are given. Finally, learn how to translate into machine learning algorithms, the problem you are trying to solve. These are the basic concepts/ideas to be mindful of but are by no means the full list of considerations one should take into account when deciding to implement machine learning into applications under development.

# Why should programmers and web developers be familiar with this technology?

Programmers and Developers can benefit immensely from having a basic understanding of how machine learning works. As discussed earlier, many tech companies like Google are incorporating machine learning into their web apps and software. As more companies start using big data, the need for machine learning will grow as well. Therefore, knowing how to use the principles of machine learning to solve business problems in addition to knowing how to program in languages like Python, Java, and C++ will become more important as Artificial Intelligence and Machine Learning become more commonplace. According to an article in the Observer with the title of *Machine Learning Is Revolutionizing Every Industry*, Machine Learning:

has already transformed industries by upending business models and, with growing maturity, will continue to be the single driver in changing how businesses across industries and nations relook at their customer engagement approach and their internal enterprise processes. In a growing global market, whether a business incorporates machine learning into its practices plays a key role in its success in the coming decades (Ramasubramanian 2016).

With the implementation of machine learning on a global scale, the impact will, in the not so distant future, be felt in far more aspects of our daily lives.

# Conclusion:

# what are the broader societal implications of this technology?

Despite the fact that an algorithm cannot be taught human qualities, they can unknowingly discriminate as they take data and classify the bits and pieces (Lebanon 2015). This was shown to be true when, in a recent study from Carnegie Mellon University, which created fake personas both male and female.  They attached the same browsing histories to the personas, and then through a third party site the male persona was shown a Google ad for senior executive positions six times more than was shown to the female persona.  Something in the algorithm picked up on the gender pay gap and then used it to discriminate against females (Datta, Tschantz and Datta 2015).

# Works Cited

Corporate. *About TensorFlow.* n.d. https://www.tensorflow.org/ (accessed February 11, 2017).

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. "Automated Experiments on Ad Privacy Settings." *Proceedings on Privacy Enhancing Technologies 2015*, 2015: 92-112.

*DeepLearning4J.* n.d. https://deeplearning4j.org/ (accessed February 11, 2017).

Gray, Alexander. *New Tech Forum.* July 7, 2015. http://www.infoworld.com/article/2943862/application-development/what-developers-need-to-know-about-machine-learning.html.

Lebanon, Guy. *The Unintended Consequences and Negative Impact of New Machine Learning Applications.* July 15, 2015. https://www.linkedin.com/pulse/unintended-consequences-negative-impact-new-machine-learning-lebanon (accessed February 11, 2017).

*Machine Learning.* February 9, 2017. https://en.wikipedia.org/wiki/Machine\_learning.

Marr, Bernard. *A Short History of Machine Learning.* February 19, 2016. http://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#66b4968c323f.

Metz, Cade. *AI Is Transforming Google Search.* February 4, 2016. https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/.

*MultiThreaded.* n.d. http://multithreaded.stitchfix.com/algorithms/ (accessed February 11, 2017).

Ramasubramanian, Guha. *Machine Learning Is Revolutionizing Every Industry.* November 28, 2016. http://observer.com/2016/11/machine-learning-is-revolutionizing-every-industry/ (accessed February 11, 2017).